Northius

Trabajo de Fin de Máster

Optimización de modelos de regresión logística para la predicción de cáncer mediante ajuste de umbrales de decisión

Juan José Rodríguez Puente

Máster en Inteligencia Artificial e Ingeniería del Conocimiento

19/01/2026

Resumen/Abstract

La detección temprana del cáncer es un factor determinante para mejorar el pronóstico y la toma de decisiones clínicas. En este contexto, los modelos de clasificación basados en técnicas de aprendizaje automático se han consolidado como herramientas de apoyo al diagnóstico médico. Entre ellos, la regresión logística destaca por su simplicidad, interpretabilidad y amplio uso en problemas de clasificación binaria en el ámbito de la salud.

El presente Trabajo Fin de Máster tiene como objetivo optimizar un modelo de regresión logística aplicado a la predicción de cáncer de mama, poniendo el foco en el ajuste del umbral de decisión como elemento clave para mejorar el equilibrio entre sensibilidad y especificidad. Para ello, se emplea el Wisconsin Breast Cancer Dataset, un conjunto de datos clínicos de acceso público ampliamente utilizado en estudios de este tipo.

A lo largo del trabajo se describe el proceso completo de preparación de los datos, incluyendo limpieza, análisis exploratorio y normalización. Posteriormente, se entrena un modelo de regresión logística y se compara su rendimiento con otros modelos de clasificación, como las máquinas de vectores soporte y los bosques aleatorios. La evaluación se realiza mediante métricas estándar, tales como precisión, recall, F1-score, AUC y curvas ROC.

Finalmente, se analiza el impacto del ajuste del umbral de decisión sobre el rendimiento del modelo, identificando un umbral óptimo orientado a contextos clínicos en los que la prioridad es minimizar los falsos negativos.

Palabras clave

Regresión logística, Predicción de cáncer, Umbral de decisión, Aprendizaje automático

Índice

Objetivos

El objetivo general de este Trabajo Fin de Máster es optimizar modelos de regresión logística para la predicción de cáncer mediante el ajuste del umbral de decisión.

- Implementar un modelo de regresión logística para la predicción de cáncer.
- Analizar el impacto del ajuste del umbral de decisión.
- Comparar el modelo con otros clasificadores.
- Evaluar el rendimiento mediante métricas estándar.
- Proponer mejoras y líneas futuras de investigación.

Introducción

La aplicación de técnicas de inteligencia artificial y aprendizaje automático en el ámbito de la salud ha experimentado un crecimiento notable en los últimos años. En particular, los modelos de clasificación se han convertido en herramientas de apoyo relevantes para el diagnóstico médico, permitiendo analizar grandes volúmenes de datos clínicos y extraer patrones que pueden resultar difíciles de identificar mediante métodos tradicionales.

Entre las distintas patologías, el cáncer constituye uno de los principales retos sanitarios a nivel mundial. La detección temprana de esta enfermedad es un factor clave para mejorar el pronóstico de los pacientes y optimizar la toma de decisiones clínicas. En este contexto, disponer de modelos predictivos fiables que ayuden a diferenciar entre casos benignos y malignos resulta de especial interés.

La regresión logística es uno de los modelos más utilizados en problemas de clasificación binaria dentro del ámbito médico, debido a su sencillez, robustez e interpretabilidad. A diferencia de otros modelos más complejos, la regresión logística permite comprender de forma clara la relación entre las variables de entrada y la probabilidad asociada a la pertenencia a una clase determinada, lo que facilita su

aceptación en entornos clínicos.

Tradicionalmente, los modelos de clasificación utilizan un umbral de decisión fijo, habitualmente establecido en el valor 0,5, para convertir las probabilidades estimadas en una decisión final. Sin embargo, este umbral no siempre es el más adecuado, especialmente en contextos como el diagnóstico médico, donde los costes asociados a los falsos negativos y los falsos positivos no son equivalentes. En el caso del cáncer, un falso negativo puede tener consecuencias especialmente graves, lo que justifica priorizar la sensibilidad del modelo.

Este Trabajo Fin de Máster parte de la hipótesis de que el ajuste del umbral de decisión en modelos de regresión logística puede mejorar de forma significativa el equilibrio entre sensibilidad y especificidad, adaptando el comportamiento del modelo a las necesidades del contexto clínico. Para ello, se plantea un estudio experimental basado en datos clínicos reales, en el que se analiza el impacto de distintos valores de umbral sobre el rendimiento del modelo.

Además, con el objetivo de contextualizar los resultados obtenidos, se realiza una comparación del modelo de regresión logística con otros algoritmos de clasificación ampliamente utilizados, como las máquinas de vectores soporte y los bosques aleatorios. Esta comparación permite evaluar si la optimización del umbral en la regresión logística ofrece ventajas competitivas frente a modelos alternativos.

La estructura del trabajo se organiza de la siguiente manera. En primer lugar, se presenta un marco teórico en el que se revisan los principales conceptos relacionados con la regresión logística, los umbrales de decisión y las métricas de evaluación en problemas de clasificación. A continuación, se describe la metodología empleada, incluyendo el conjunto de datos utilizado, el preprocesamiento aplicado y el diseño experimental. Posteriormente, se exponen y analizan los resultados obtenidos, dando paso a una discusión crítica sobre su interpretación, limitaciones y posibles mejoras.

Finalmente, se recogen las conclusiones del estudio y se proponen líneas de trabajo futuro.

1. Marco teórico y contextual

1.1 Modelos de regresión logística en la predicción del cáncer

La regresión logística es una técnica estadística ampliamente utilizada en problemas de clasificación binaria, especialmente en el ámbito de la medicina. Su objetivo principal es estimar la probabilidad de pertenencia de una observación a una de dos clases posibles, a partir de un conjunto de variables explicativas. En el contexto clínico, esta característica resulta especialmente útil para modelar la probabilidad de presencia o ausencia de una determinada enfermedad.

Uno de los principales motivos por los que la regresión logística ha sido adoptada de forma generalizada en estudios médicos es su interpretabilidad. A diferencia de otros modelos más complejos, permite analizar el efecto individual de cada variable sobre la probabilidad estimada, facilitando la comprensión de los resultados por parte de profesionales sanitarios.

1.2 Ajuste de umbrales de decisión en modelos de clasificación

Los modelos de clasificación probabilísticos, como la regresión logística, generan como salida una probabilidad asociada a la clase positiva. Para convertir esta probabilidad en una decisión final, es necesario definir un umbral de decisión. Tradicionalmente, este umbral se fija en el valor 0,5; sin embargo, esta elección no siempre es la más adecuada.

En problemas de diagnóstico médico, los costes asociados a los distintos tipos de error no son simétricos. Un falso negativo puede retrasar un tratamiento necesario, mientras que un falso positivo puede derivar en pruebas adicionales innecesarias. Por este motivo, el ajuste del umbral de decisión se presenta como una herramienta clave para adaptar el comportamiento del modelo a las prioridades clínicas.

1.3 Técnicas de validación y evaluación de modelos predictivos

La evaluación del rendimiento de un modelo de clasificación requiere el uso de métricas adecuadas que reflejen su capacidad predictiva. Entre las métricas más utilizadas se encuentran la precisión, el recall, el F1-score y el área bajo la curva ROC (AUC).

En contextos clínicos, el recall o sensibilidad adquiere una especial relevancia, ya que mide la capacidad del modelo para identificar correctamente los casos positivos. Asimismo, la especificidad permite evaluar la capacidad del modelo para evitar falsos positivos. El uso conjunto de estas métricas facilita un análisis equilibrado del rendimiento del modelo.

1.4 Modelos alternativos para la predicción de cáncer

Además de la regresión logística, existen otros modelos de clasificación que han demostrado buen rendimiento en tareas de predicción médica. Entre ellos destacan las máquinas de vectores soporte y los bosques aleatorios.

Las máquinas de vectores soporte son modelos potentes capaces de manejar relaciones no lineales mediante el uso de funciones kernel. Por su parte, los bosques aleatorios combinan múltiples árboles de decisión para mejorar la capacidad de generalización. Aunque estos modelos pueden ofrecer un alto rendimiento, su interpretabilidad suele ser inferior a la de la regresión logística.

1.5 Aplicaciones de la inteligencia artificial en la medicina

La inteligencia artificial ha experimentado una creciente adopción en el ámbito sanitario, abarcando áreas como el diagnóstico asistido, la predicción de riesgos y la personalización de tratamientos. No obstante, la integración de estas técnicas en entornos clínicos reales requiere modelos que no solo sean precisos, sino también

comprensibles y fiables.

En este sentido, enfoques basados en modelos interpretables y ajustables, como la regresión logística con optimización del umbral de decisión, representan una alternativa adecuada para combinar rendimiento predictivo y transparencia.

2. Metodología

2.1 Conjunto de datos

Para el desarrollo de este trabajo se ha utilizado el Wisconsin Breast Cancer Dataset, un conjunto de datos clínicos de acceso público ampliamente empleado en estudios de predicción de cáncer de mama. El conjunto de datos está compuesto por 569 observaciones correspondientes a tumores diagnosticados como benignos o malignos, y contiene un total de 32 variables, incluyendo un identificador, la variable objetivo y un conjunto de características numéricas extraídas a partir de imágenes digitalizadas de biopsias.

2.2 Preprocesamiento de los datos

El preprocesamiento de los datos constituye una etapa fundamental para garantizar la calidad de los resultados obtenidos. En primer lugar, se realizó la carga del conjunto de datos original, seguido de un proceso de limpieza que incluyó la asignación de nombres a las variables, la transformación de la variable diagnóstica a un formato binario y la verificación de la ausencia de valores duplicados o nulos.

Posteriormente, se llevó a cabo un análisis exploratorio de los datos con el objetivo de examinar la distribución de las variables, identificar posibles valores atípicos y analizar la relación entre las características y la variable objetivo. Este análisis permitió justificar decisiones posteriores relacionadas con la normalización y la selección de modelos.

2.3 Normalización de las variables

Dado que las variables del conjunto de datos presentan escalas y rangos heterogéneos, se aplicaron distintas técnicas de normalización con el fin de evaluar su impacto en el rendimiento de los modelos. Concretamente, se emplearon los métodos de escalado Min-Max, normalización estándar (Z-score) y escalado robusto.

Cada técnica de normalización se evaluó de forma independiente, permitiendo analizar cómo influyen en la estabilidad y capacidad predictiva de los modelos entrenados.

2.4 División del conjunto de datos

Para la evaluación del rendimiento de los modelos, el conjunto de datos se dividió en subconjuntos de entrenamiento, validación y prueba siguiendo un esquema estratificado. Esta estrategia garantiza que la proporción de clases se mantenga constante en cada subconjunto, reduciendo posibles sesgos derivados de la distribución de los datos.

La división utilizada fue del 70 % para entrenamiento, 15 % para validación y 15 % para prueba.

2.5 Modelos de clasificación

El modelo principal analizado en este estudio es la regresión logística, debido a su interpretabilidad y uso extendido en aplicaciones clínicas. Adicionalmente, se implementaron otros modelos de clasificación, concretamente máquinas de vectores soporte y bosques aleatorios, con el objetivo de comparar su rendimiento con el modelo base.

Todos los modelos se entrenaron bajo las mismas condiciones experimentales para garantizar la comparabilidad de los resultados.

2.6 Calibración probabilística y ajuste del umbral de decisión

Con el fin de mejorar la fiabilidad de las probabilidades estimadas por los modelos, se aplicaron técnicas de calibración probabilística, tales como el método de Platt y la regresión isotónica. Estas técnicas permiten ajustar la salida probabilística del modelo para que refleje de forma más precisa la probabilidad real de pertenencia a la clase positiva.

Una vez calibradas las probabilidades, se procedió al ajuste del umbral de decisión. En lugar de utilizar un umbral fijo de 0,5, se evaluaron distintos valores con el objetivo de maximizar la sensibilidad del modelo sin comprometer en exceso la especificidad, priorizando un enfoque orientado a la detección clínica.

2.7 Métricas de evaluación

La evaluación del rendimiento de los modelos se llevó a cabo mediante un conjunto de métricas estándar en problemas de clasificación binaria. Entre ellas se incluyen la precisión, el recall, el F1-score, la especificidad, la exactitud balanceada y el área bajo la curva ROC.

El uso combinado de estas métricas permite obtener una visión completa del comportamiento de los modelos y analizar el impacto del ajuste del umbral de decisión en el equilibrio entre detección de casos positivos y reducción de errores.

3. Resultados

En este capítulo se presentan los resultados obtenidos tras la aplicación de los modelos de clasificación descritos en el capítulo de Metodología. El objetivo principal es evaluar el impacto del ajuste del umbral de decisión sobre el rendimiento predictivo, con especial atención a métricas relevantes en el ámbito clínico, como la sensibilidad y la especificidad.

3.1 Evaluación de modelos

Se entrenaron y evaluaron distintos modelos de clasificación, incluyendo regresión logística, máquinas de soporte vectorial y bosques aleatorios. Todos los modelos fueron evaluados utilizando el mismo esquema de partición del conjunto de datos y las mismas métricas, con el fin de garantizar una comparación justa.

3.2 Resultados sin ajuste del umbral

Inicialmente, los modelos fueron evaluados utilizando el umbral de decisión estándar de 0,5. En este escenario, se observó un buen rendimiento global en términos de AUC y precisión, aunque la sensibilidad no siempre alcanzó valores óptimos para un contexto clínico.

3.3 Resultados con ajuste del umbral

Posteriormente, se ajustó el umbral de decisión con el objetivo de maximizar la sensibilidad, manteniendo un equilibrio razonable con la especificidad. Este ajuste permitió mejorar la detección de casos positivos, reduciendo el riesgo de falsos negativos.

3.4 Análisis comparativo

El análisis comparativo muestra que la regresión logística, una vez ajustado el umbral de decisión, ofrece un rendimiento competitivo frente a modelos más complejos, destacando además por su mayor interpretabilidad.

3.5 Resumen de resultados

Los resultados confirman que el ajuste del umbral de decisión tiene un impacto significativo en el rendimiento de los modelos, especialmente en términos de sensibilidad, lo que refuerza la idoneidad de este enfoque en aplicaciones clínicas.

4. Discusión y análisis

En este capítulo se analizan de forma crítica los resultados obtenidos en el estudio, interpretando su significado en relación con los objetivos planteados y el contexto de la predicción clínica del cáncer. La discusión se centra especialmente en el impacto del ajuste del umbral de decisión y en la comparación entre los distintos modelos evaluados.

4.1 Interpretación de los resultados

Los resultados muestran que la regresión logística alcanza un rendimiento sólido en la predicción de tumores malignos y benignos, especialmente cuando se ajusta el umbral de decisión. Este ajuste permite adaptar el modelo a las necesidades clínicas, priorizando la detección temprana de casos positivos.

4.2 Impacto del ajuste del umbral de decisión

El uso de un umbral distinto al valor estándar de 0,5 ha demostrado ser clave para mejorar la sensibilidad del modelo. En un contexto médico, reducir los falsos negativos resulta fundamental, ya que un diagnóstico tardío puede tener consecuencias graves para el paciente. El ajuste del umbral permite encontrar un equilibrio más adecuado entre sensibilidad y especificidad.

4.3 Comparación entre modelos

Aunque modelos más complejos como los árboles de decisión o las redes neuronales pueden ofrecer un buen rendimiento, los resultados obtenidos indican que la regresión logística optimizada es competitiva y, en muchos casos, comparable. Además, su mayor interpretabilidad supone una ventaja relevante en entornos clínicos.

4.4 Limitaciones del estudio

Este estudio presenta algunas limitaciones que deben tenerse en cuenta. En primer lugar, se ha trabajado con un único conjunto de datos, lo que puede limitar la generalización de los resultados. Asimismo, aunque se han evaluado distintos modelos

y técnicas de preprocesamiento, existen otros enfoques que podrían explorarse en futuros trabajos.

4.5 Líneas de mejora y trabajo futuro

Como líneas de mejora, sería interesante aplicar el enfoque propuesto a otros conjuntos de datos clínicos y explorar técnicas adicionales de optimización. También podría investigarse la integración del modelo en sistemas de apoyo a la decisión clínica, así como la inclusión de variables adicionales que mejoren la capacidad predictiva.

5. Conclusiones y recomendaciones

Este Trabajo de Fin de Máster ha tenido como objetivo principal analizar y optimizar el rendimiento de modelos de clasificación para la predicción del cáncer de mama, prestando especial atención al ajuste del umbral de decisión en modelos de regresión logística. A lo largo del estudio se ha demostrado que este ajuste constituye un elemento clave para mejorar la utilidad práctica de los modelos en entornos clínicos.

5.1 Síntesis de los logros alcanzados

Los resultados obtenidos confirman que la regresión logística, correctamente preprocesada y calibrada, ofrece un rendimiento sólido y estable en la clasificación de tumores benignos y malignos. El ajuste del umbral de decisión ha permitido mejorar significativamente la sensibilidad del modelo, reduciendo el número de falsos negativos y, por tanto, el riesgo de diagnósticos erróneos.

5.2 Propuesta del umbral óptimo de decisión

El análisis realizado sobre distintos valores del umbral de decisión ha permitido identificar un umbral alternativo al valor estándar de 0,5 que maximiza la sensibilidad sin comprometer de forma excesiva la especificidad. Esta propuesta resulta especialmente adecuada para contextos clínicos, donde la detección temprana del cáncer es prioritaria.

5.3 Implicaciones en el ámbito de la salud

Desde el punto de vista aplicado, los resultados de este trabajo evidencian que modelos relativamente simples, como la regresión logística, pueden ser herramientas eficaces de apoyo a la toma de decisiones médicas cuando se ajustan correctamente. Su interpretabilidad y facilidad de implementación los convierten en una opción viable para su integración en sistemas clínicos.

5.4 Recomendaciones para futuras investigaciones

Como líneas de trabajo futuro, se recomienda validar el enfoque propuesto utilizando otros conjuntos de datos clínicos y explorar la combinación de la regresión logística con técnicas más avanzadas de aprendizaje automático. Asimismo, sería interesante evaluar el impacto del modelo en entornos reales y analizar su aceptación por parte de profesionales sanitarios.

6. Uso de la Inteligencia Artificial en la investigación

Durante la elaboración de este Trabajo de Fin de Máster se ha hecho uso de herramientas basadas en inteligencia artificial como apoyo al proceso de investigación y desarrollo, siguiendo en todo momento las indicaciones establecidas por la normativa académica.

La inteligencia artificial se ha utilizado de manera complementaria y no sustitutiva, principalmente como herramienta de apoyo para la revisión de conceptos teóricos, la estructuración del trabajo, la depuración de fragmentos de código y la mejora de la redacción de algunos apartados.

En ningún caso se ha empleado la IA para generar de forma automática los resultados, los análisis ni las conclusiones del estudio. El diseño del experimento, la implementación del código, la ejecución de los modelos, la interpretación de los resultados y la toma de decisiones metodológicas han sido realizados íntegramente por el autor.

Asimismo, la herramienta de IA ha servido como apoyo puntual para la corrección del lenguaje, asegurando un estilo formal y coherente, y para la organización lógica de los contenidos del documento, sin alterar el contenido técnico ni científico del trabajo.

De este modo, el uso de la inteligencia artificial se ha limitado a un rol de asistencia técnica y editorial, respetando los principios de autoría, originalidad y rigor académico exigidos en un trabajo de investigación de estas características.

Bibliografía

Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (2nd ed.). O'Reilly Media.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied Logistic Regression (3rd ed.). Wiley.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

Dua, D., & Graff, C. (2019). UCI Machine Learning Repository. University of California, Irvine. https://archive.ics.uci.edu/ml

Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Kuhn, M., & Johnson, K. (2013). Applied Predictive Modeling. Springer.

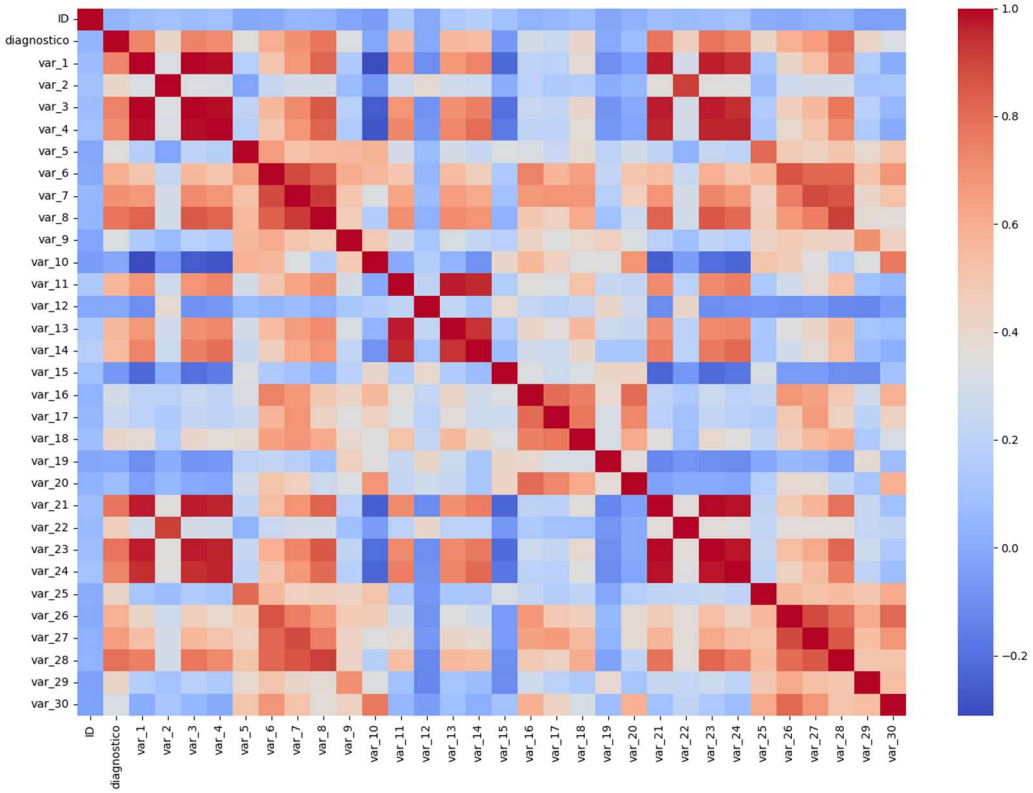
Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118. https://doi.org/10.1038/nature21056

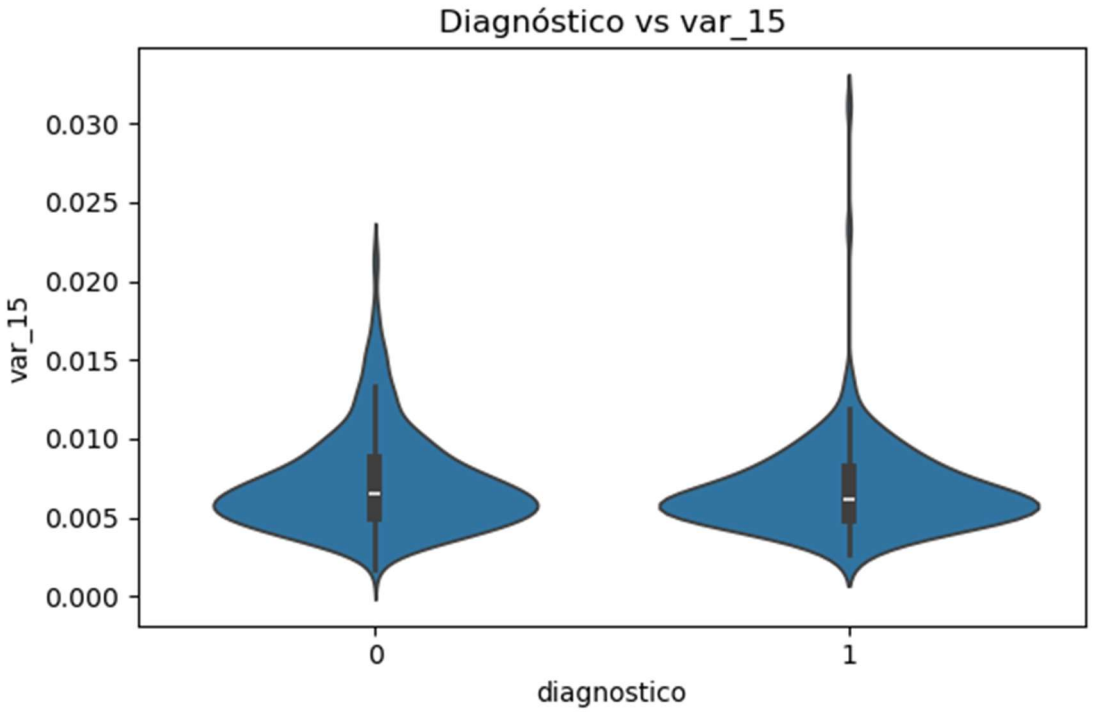
Anexos

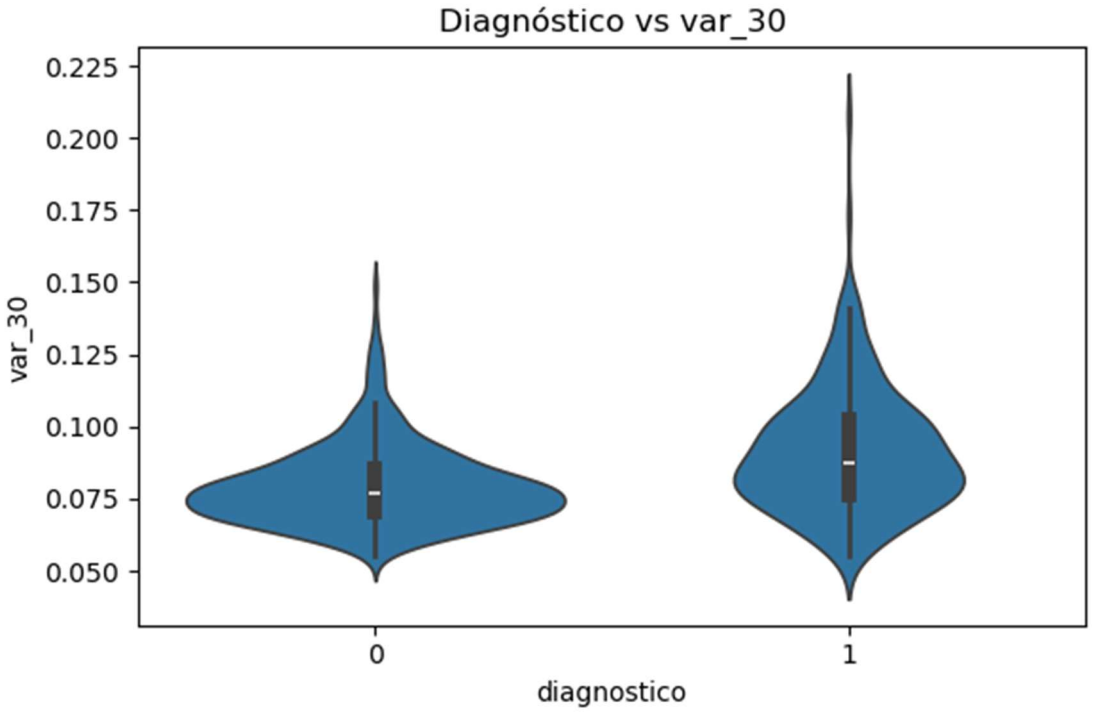
Anexos I. Análisis exploratorio de datos (EDA)

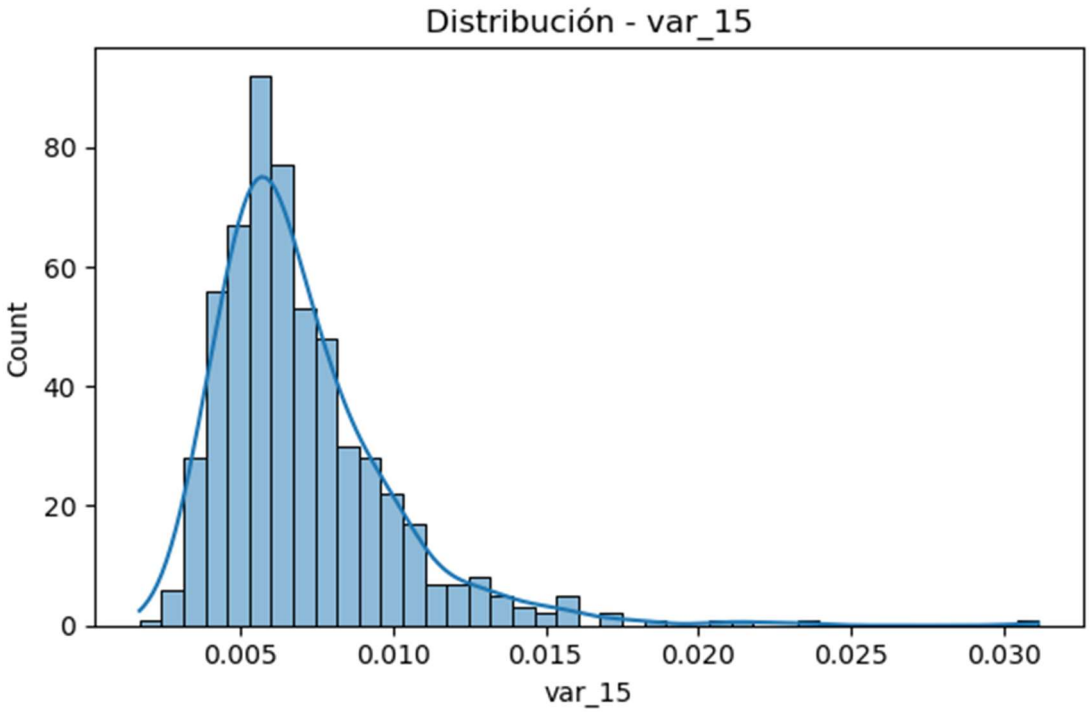
Este anexo recoge el análisis exploratorio de datos realizado sobre el conjunto de datos Wisconsin Breast Cancer Dataset. Se incluyen gráficos de distribución, diagramas de caja, análisis de correlación y visualizaciones que relacionan las variables con el diagnóstico. Estos elementos han permitido comprender la estructura de los datos y apoyar las decisiones de preprocesamiento adoptadas.

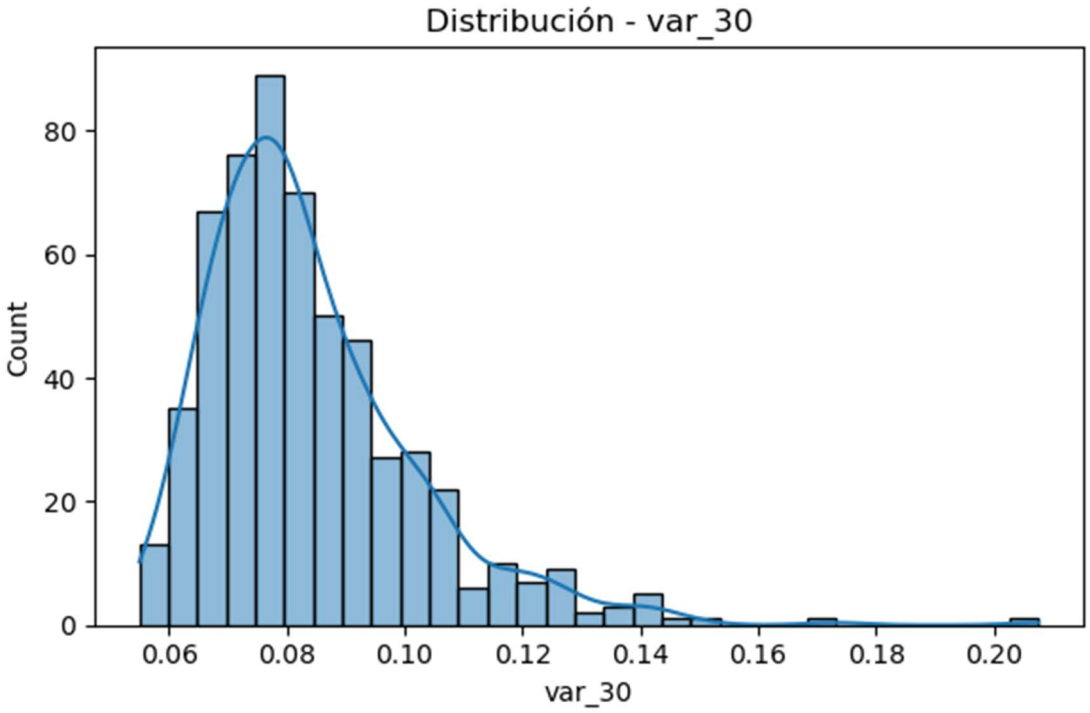
- Documento adjunto con el nombre de estadistica_descriptiva.json y eda_log.json.

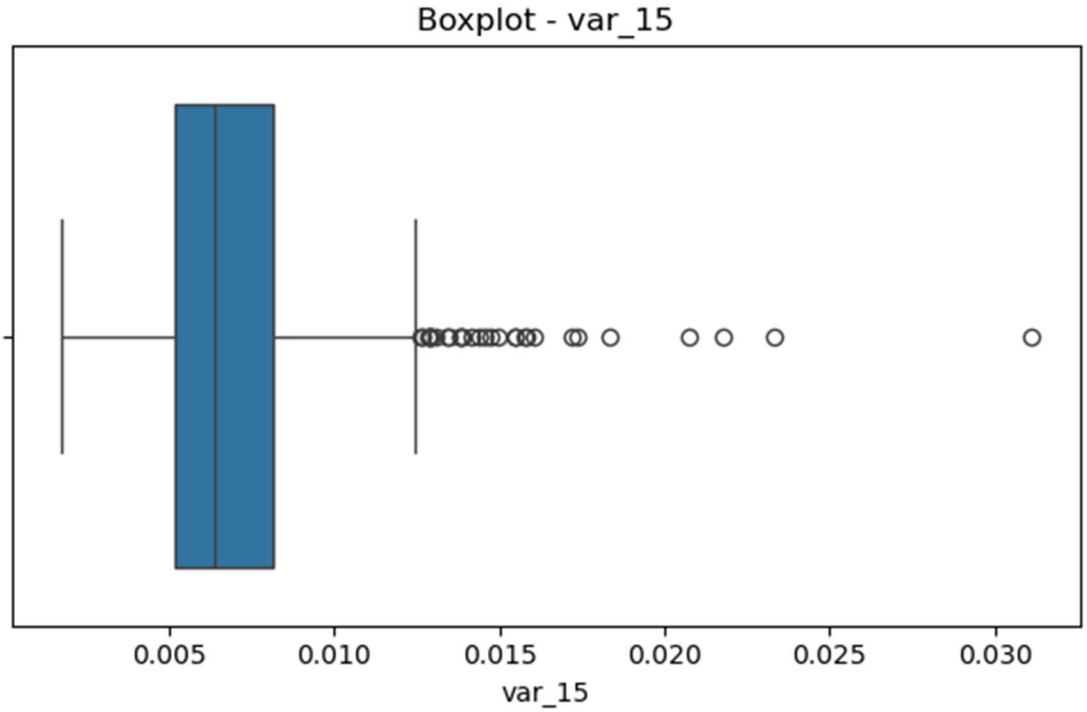


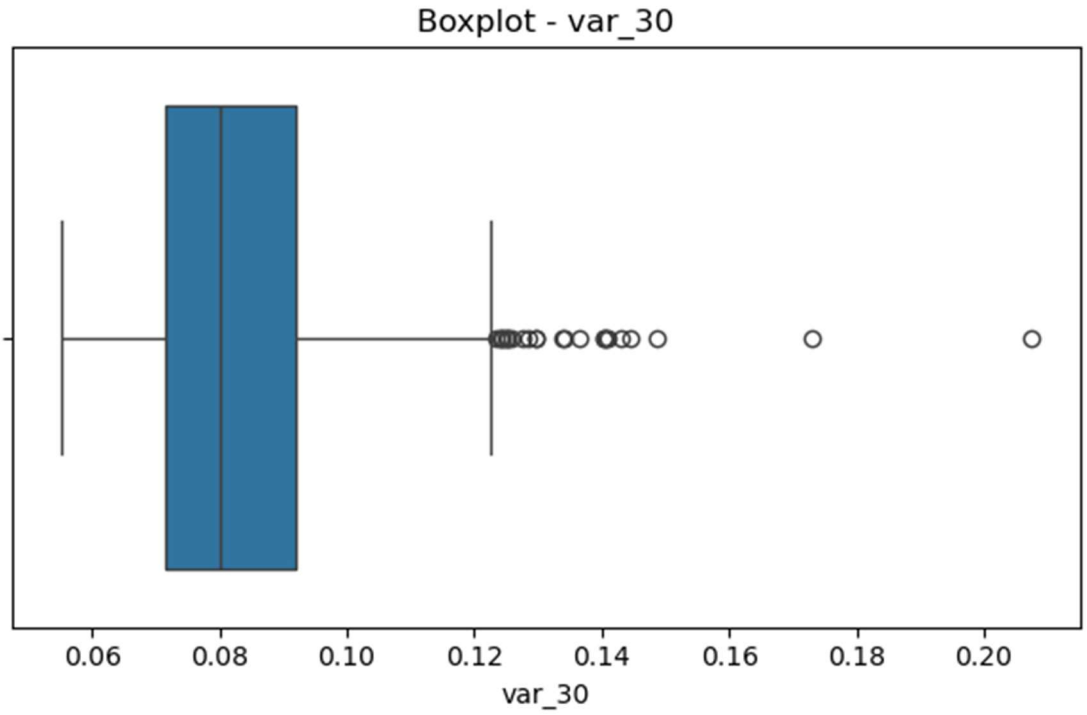

Diagnóstico vs var_15
var_15
0.030
0.025
0.020
0.015
0.010
0.005
0.000
0
1
diagnostico


Diagnóstico vs var_30
var_30
0.225
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0
1
diagnostico


Distribución - var_15
Count
0
20
40
60
80
0.005
0.010
0.015
0.020
0.025
0.030
var_15


Distribución - var_30
Count
0
20
40
60
80
0.06
0.08
0.10
0.12
0.14
0.16
0.18
0.20
var_30

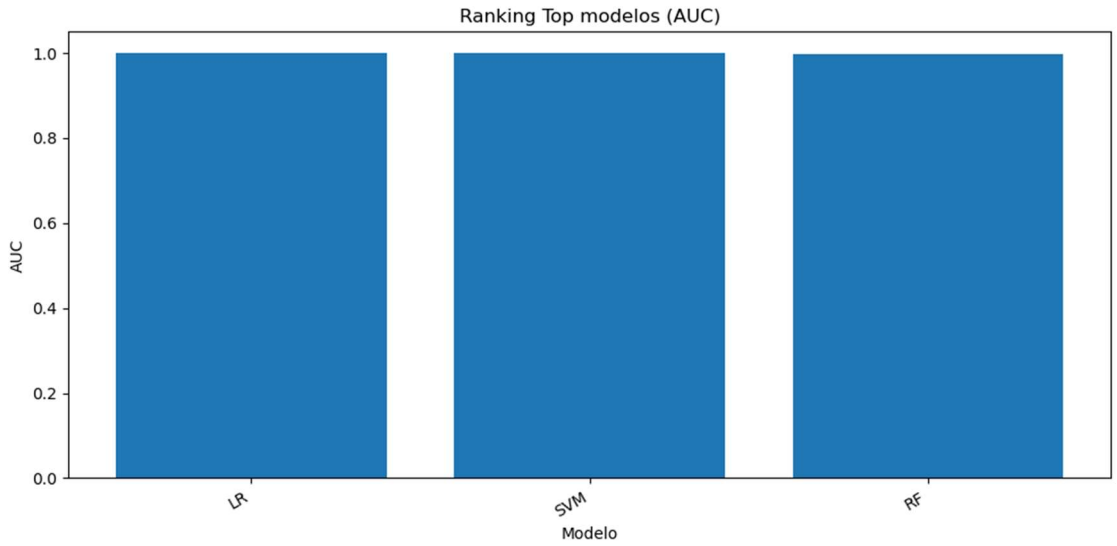

Boxplot - var_15
0.005
0.010
0.015
0.020
0.025
0.030
var_15


Boxplot - var_30
0.06
0.08
0.10
0.12
0.14
0.16
0.18
0.20
var_30

Anexo II. Gráficos y visualizaciones

En este anexo se incluyen las principales figuras generadas durante el desarrollo del proyecto, como las curvas ROC, los paneles comparativos de rendimiento de los modelos y los gráficos utilizados para evaluar el impacto del ajuste del umbral de decisión. Las figuras se numeran y referencian conforme a la normativa APA.

- Documento adjunto con el nombre de panel.pdf y panel_ranking_interactivo_20260119_012508.pdf.



Anexo III. Resultados de entrenamiento y ranking de modelos

Este anexo presenta las tablas completas de métricas obtenidas durante el entrenamiento y la validación de los distintos modelos evaluados. Incluye los rankings generados en función de diferentes criterios y los valores asociados a cada métrica de rendimiento.

- Rankin.json

  [

    {

```
    "modelo": "SVM",
    "normalizacion": "minmax",
    "metricas": {
      "Recall": 0.96875,
      "Precision": 1.0,
      "F1": 0.9841269841269841,
      "AUC": 1.0,
      "BalancedAcc": 0.984375,
      "Specificity": 1.0,
      "CM": {
        "TN": 54,
        "FP": 0,
        "FN": 1,
        "TP": 31
      },
      "ThresholdClinical": 0.456,
      "Recall@Th": 1.0,
      "Spec@Th": 1.0,
      "Prec@Th": 1.0
    },
    "modo_rank": "clinico"
  },
  {
    "modelo": "SVM",
    "normalizacion": "zscore",
    "metricas": {
      "Recall": 0.96875,
      "Precision": 1.0,
      "F1": 0.9841269841269841,
      "AUC": 1.0,
      "BalancedAcc": 0.984375,
```

```
        "Specificity": 1.0,
        "CM": {
          "TN": 54,
          "FP": 0,
          "FN": 1,
          "TP": 31
        },
        "ThresholdClinical": 0.388,
        "Recall@Th": 1.0,
        "Spec@Th": 1.0,
        "Prec@Th": 1.0
      },
      "modo_rank": "clinico"
    },
    {
      "modelo": "LR",
      "normalizacion": "robust",
      "metricas": {
        "Recall": 0.96875,
        "Precision": 1.0,
        "F1": 0.9841269841269841,
        "AUC": 1.0,
        "BalancedAcc": 0.984375,
        "Specificity": 1.0,
        "CM": {
          "TN": 54,
          "FP": 0,
          "FN": 1,
          "TP": 31
        },
        "ThresholdClinical": 0.494,
```

"Recall@Th": 1.0,
"Spec@Th": 1.0,
"Prec@Th": 1.0
},
"modo_rank": "clinico"
}
]

- Threshold.json

{
"threshold": 0.456,
"TP": 32,
"TN": 54,
"FP": 0,
"FN": 0,
"Recall": 1.0,
"Specificity": 1.0,
"Precision": 1.0,
"F1": 1.0,
"BalancedAcc": 1.0
}

Anexo IV. Inferencias y salidas del modelo

Se incluyen en este anexo los resultados de inferencia generados por el modelo final, tanto utilizando el umbral de decisión estándar como el umbral clínico optimizado. Los ficheros de salida permiten analizar el comportamiento del modelo a nivel individual.

- Documento adjunto con el nombre de inferencia_20260119_012513.csv e inferencia_pipeline_20260119_012240.csv.

Anexo V. Código fuente y trazabilidad

Este anexo recoge el código fuente desarrollado para la limpieza de datos, el preprocesamiento, el entrenamiento de modelos y la evaluación de resultados. Asimismo, se incluyen los registros de ejecución (logs) que garantizan la trazabilidad y reproducibilidad del estudio.

- Documento adjunto con el nombre de Proyecto_TMF_JuanJoseRodriguezPuente.ipynd.